

Moving Object Detection Using Sparse Approximation and Sparse Coding Migration

Shufang Li^{1,2}, Zhengping Hu^{1,*} and Mengyao Zhao¹

¹ School of Information and Engineering, Yanshan University
Qinhuangdao, Hebei 066004 - China
[e-mail: hzp_ysu@163.com]

² Department of Information Engineering, Hebei University of Environmental Engineering
Qinhuangdao, Hebei 066004 - China
[e-mail: lishufang116003@sina.com]

*Corresponding author: Zhengping Hu

*Received July 26, 2019; revised October 24, 2019; revised January 15, 2020; accepted February 13, 2020;
published May 31, 2020*

Abstract

In order to meet the requirements of background change, illumination variation, moving shadow interference and high accuracy in object detection of moving camera, and strive for real-time and high efficiency, this paper presents an object detection algorithm based on sparse approximation recursion and sparse coding migration in subspace. First, low-rank sparse decomposition is used to reduce the dimension of the data. Combining with dictionary sparse representation, the computational model is established by the recursive formula of sparse approximation with the video sequences taken as subspace sets. And the moving object is calculated by the background difference method, which effectively reduces the computational complexity and running time. According to the idea of sparse coding migration, the above operations are carried out in the down-sampling space to further reduce the requirements of computational complexity and memory storage, and this will be adapt to multi-scale target objects and overcome the impact of large anomaly areas. Finally, experiments are carried out on VDAO datasets containing 59 sets of videos. The experimental results show that the algorithm can detect moving object effectively in the moving camera with uniform speed, not only in terms of low computational complexity but also in terms of low storage requirements, so that our proposed algorithm is suitable for detection systems with high real-time requirements.

Keywords: Object detection, sparse approximation, sparse coding migration, background modeling, moving camera

1. Introduction

In the field of computer vision and pattern recognition, moving object detection is a very important and active research direction. At present, related detection has been widely applied in many fields, such as industrial monitoring, traffic control, robot navigation, action recognition and intelligent monitoring.

In order to solve the problems of background change, illumination change and moving shadow interference, relative researchers have done a lot of research on the algorithm and model to detect the object in the camera accurately, real-time and efficiently. With the development of compressed sensing and sparse representation theory in the early stage, sparse representation and dictionary learning were applied to moving object detection, and achieved some results. Ref. [1] used the sparse representation method to extract the structural features, and attained the effective sparse representation of the image. Ref. [2] proposed robust sparse representation model, in which the least square error is used instead of the sparse error recently. Replacing the original image data representation with sparse approximation, this can substantially reduce the storage requirements, improve the operation speed, and reduce the processing cost, but the sparse coefficient has no spatial and temporal correlation. Zhao et al. used training samples to update dictionary atoms, and reconstructed background model using dictionary atoms and sparse coefficients [3]. Because the model does not update the background model with the nearest frame images, the adaptive ability of the background model is not strong enough. Video sequences can be approximated by data dictionary, but data dictionary is relatively large in the process of sparse decomposition and target reconstruction. And some important data will be lost at the same time. Therefore, some researchers propose to reduce the dimensionality of high-dimensional data by subspace learning model. Oliver et al. first used Principal Component Analysis (PCA) to construct background model, which greatly reduced the dimensions of video data [4]. Classical PCA can achieve dimensionality reduction well when the noise is small. However, the method will be affected when the noise is large, even if the noise only affects a part of the matrix. Then Robust Principal Component Analysis (RPCA) was proposed to solve this problem. Subsequently, Javed et al. created the background model by using an improved RPCA to generate rank matrix from a set of matrices [5]. For moving camera, it is necessary to transform two-dimensional parameter into the model to make up the background variation caused by camera. Hu et al. proposed using tensor kernel norm to develop background spatio-temporal redundancy, and utilizing fusion sparse regularizer to perform adaptive constraints on spatio-temporal smoothing of foreground [6]. Thomas et al. completed the low-rank representation of abnormal object by taking all the frames of the video as a joint subspace and calculating the sparse residual of the reference video in low-speed motion of the camera [7]. Although low rank and sparse matrix decomposition has prominent advantages in the detection of object in moving camera, its computational complexity is relatively high and its real-time performance is poor. So Qin et al. tried to achieve fast and robust moving object detection method by graph cutting and code migration [8]. At the same time, Ref. [9] proposes a benchmark to test the target tracking method of high frame rate dataset in camera, which can effectively assess the performance of the method from two aspects of speed and accuracy. Recently, de Carvalho et al. proposed a multi-resolution method to detect objects of different sizes from moving camera in real time and effectively [10].

On the basis of ensuring the accuracy, this paper focuses on the real-time problem, and proposes the object detection algorithm based on sparse approximation recursion and sparse coding migration. We take the continuous background video sequence as a series of subspace sets. On the basis of low rank and sparse matrix decomposition, the sparse approximation recursion is used to replace conventional updating about background model, so that the complexity and computation of the model will be reduced. Then the background subtraction is used to calculate the moving object. According to the idea of sparse coding migration, the target image video and background video are calculated in the down-sampling space, which further reduces the computational complexity, lowers the storage requirements, and overcomes the impact of abnormal areas. Through the above two improvements, the calculation of the background modeling can meet the real-time requirement of the system.

2. Related Work

2.1 Background Difference Modeling

Background difference method is developed on the basis of inter-frame difference, which effectively eliminates the problems of holes and false targets. First, a background model is established and compared with each frame image in the video continuously. The region similar to the background in the image is considered as the background, and the region that can not match is called the foreground. And then the background model is updated with the image information for the next detection. The flow chart is shown in [Fig. 1](#).

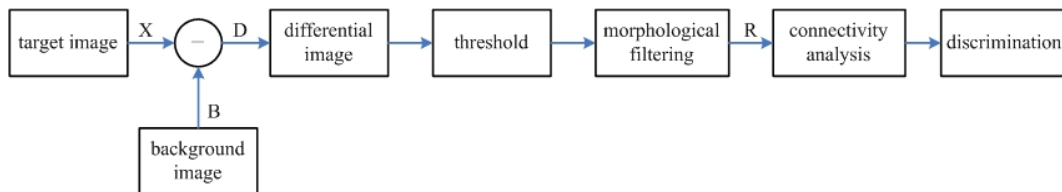


Fig. 1. Background difference modeling

The specific processing contains pretreatment, background modeling, foreground detection and post-processing. Background modeling mainly includes initial background modeling, extracting feature points from the current frame, establishing association between feature matching and corresponding background, calculating conversion matrix between them, dividing foreground and background pixels, and updating background model. For moving camera, background modeling requires motion compensation for the uncertain relationship between the background model and the coordinates of the pixels caused by camera movement [11, 12]. Recently, Zhou and Maskell applied background subtraction and motion compensation to solve the problem about object detection in urban video captured from aerial moving camera [13]. Despite the high complexity of the algorithm, it can still detect moving objects effectively in the case of large parallax in video sequences. Meanwhile, Gong et al. proposed background subtraction model based dictionary for camera object detection on moving vehicle, in which full-color information was used to detect object [14]. Background difference modeling has better performance for flat camera motion and specific PTZ situation, and it is widely used in real-time applications for its moderate complexity.

2.2 Sparse Representation and Dictionary Learning

The sparse representation of matrices can be divided into the sparsity of matrix elements and the sparsity of matrix singular values. Generally speaking, matrix sparse representation refers to the sparsity of matrix elements. The basic idea is to assume that natural signals can be compressed or expressed by predefined linear combinations of atoms.

Let original signal be $x \in R^M$, data dictionary be $D = [d_1, d_2, \dots, d_L] \in R^{M \times L}$, and $d_i \in R^M$ ($i = 1, 2, \dots, L$) be the atoms of the dictionary. Each signal can be expressed by a linear combination of a set of dictionary atoms:

$$x = D\alpha = \sum_{i=1}^L d_i \alpha_i \quad (1)$$

where α is the sparse representation coefficient.

The coefficient expressed by dictionary D is sparse, namely, just a few atoms are required to unique linear representation of image. The problem of image sparse representation can be expressed as [15]:

$$\alpha = \arg \min_{\alpha} \|\alpha\|_0, \quad s.t. \|x - D\alpha\|_2 \leq \varepsilon \quad (2)$$

where $\|\alpha\|_0$ is l_0 norm, ε is sparse representation error, and error constraint $\|x - D\alpha\|_2 \leq \varepsilon$ ensures that sparse α get close to x . When the signal and dictionary satisfy certain conditions, l_0 norm can be converted to the convex optimization problem of l_1 norm to solve [16]. Eq. (2) can be transformed into:

$$\alpha = \arg \min_{\alpha} \|\alpha\|_1, \quad s.t. \|x - D\alpha\|_2 \leq \varepsilon \quad (3)$$

By Lagrange multiplier method, the solution is as follow:

$$\alpha = \arg \min_{\alpha} \|x - D\alpha\|_2^2 + \mu \|\alpha\|_1 \quad (4)$$

where $\|\alpha\|_1 = \sum_i \alpha_i$, and μ is penalty factor. Given the appropriate condition, the solution of

Eq. (3) and Eq. (4) is equivalent, and its solution can be achieved by sparse decomposition algorithm.

For sparse representation of image, dictionary construction is very important. K-SVD is a classical dictionary training algorithm. According to the principle of minimum error, the error term can be decomposed by SVD. We can choose the decomposition term with the smallest error as the updated dictionary atom and the corresponding atom coefficient. And the optimized solution is obtained through continuous iteration. Let $D \in R^{n \times k}$, $y \in R^n$ and $x \in R^k$ represent dictionary, training signal and coefficient vector of signal in sparse representation respectively. $Y = \{y_i\}_{i=1}^N$ is a set of N training signals. And $X = \{x_i\}_{i=1}^N$ is a

set of solution vectors for Y . The construction of K-SVD learning dictionary can be expressed by sparse constraint:

$$D = \arg \min_{D, X} \|Y - DX\|_2^2, \quad s.t. \begin{cases} \forall i, \|x_i\|_0 \leq T_0 \\ \forall j, \|d_j\|_2 = 1 \end{cases} \quad (5)$$

D and X are the reconfiguration goals. T_0 is the maximum number of non-zero components in the sparse representation coefficient X . And the scale ambiguity of principle is avoided through l_2 norm constraint of each atom d_j in D .

Video target tracking algorithm based on sparse representation is essentially a sparse approximation problem under redundant dictionary, which can find out the structure and mode hidden in the input data more effectively. Sparse representation algorithm based on dictionary learning can solve the shortcomings of traditional moving object detection algorithm. Through the building of sparse model of video sequence, redundant information can be effectively removed and the structural features of inter-frame images can be reflected. Nakahata et al. used a two-stage dictionary learning process and improved spatiotemporal features to complete anomaly detection in moving camera [17]. The establishment of double dictionaries can solve the shortcomings of single dictionary and take full account of the influence of illumination uniformity.

In order to improve the ability of dealing with complex background and the real-time response of the model, Ref. [18] constructed a dictionary by using both target and background features, in which the dictionary itself is discriminative and the sparse representation contain discriminated information, so that the algorithm had a strong ability to distinguish between target and background.

2.3 Low Rank and Sparse Matrix Decomposition

In the problem of moving object detection, the video often contains multiple images, and the backgrounds between consecutive frames usually have great similarity, so there are some correlations and redundancies in the video data. Therefore, subspace learning model can be used to reduce the dimension of video data and obtain low-dimensional background model, so as to detect moving objects efficiently. When detecting moving object, the video data to be detected can also be expressed in matrix form, and then the subspace learning model is used to realize the task of moving object detection. Early Zhou et al. put forward to detect moving object through checking the low rank of continuous outliers [19]. Subsequently, Oliver et al. proposed a feature space model for each background pixel based on RPC algorithm [4]. Later RPCA algorithm makes up for the problem that PCA algorithm is difficult to guarantee robustness to missing data. The moving object detection method based on RPCA assumes that the observation matrix consists of two parts, one is the low rank matrix corresponding to the relatively stable background, the other is the sparse matrix corresponding to the noise and the foreground moving object. On the basis of RPCA, Javed et al. used motion perception regularization of low rank component graph to detect object, and added information such as optical flow, intra-frame and inter-frame graph to background modeling [5]. For slow-moving camera, the low-rank dual-sparse model can be used to represent the continuous frames of a given background video [20]. Jardim et al. used subspace restoration and sparse decomposition to complete anomaly detection, and explored the low-rank similarity between reference video and target video as well as the sparseness of differences between target video

soon afterwards. Recent study proposed introducing spatial information of object into object detection algorithm in subspace learning architecture [10]. Although low rank and sparse decomposition has prominent advantages in the detection of dynamic objects in moving camera, it has relatively high computational complexity and poor real-time performance.

2.4 Real-time Problem of Object Detection

The real-time problem is mainly due to the need to design an effective detection algorithm in moving object detection to improve the accuracy and reduce the error. In addition, the detection algorithm is required to have high detection efficiency because detection needs to deal with a large number of video sequences. Ref. [21] can effectively detect abnormal object in complex background, but the computational complexity of algorithm is too large to meet the real-time requirements. The complexity of the algorithm increases significantly with the increase of video size, therefore the Ref. [21] can only process low-pixel video for a very short time, that is, a video clip with a resolution of 320×180 pixels per frame and a total of 70 frames. In order to reduce the operation time, some inherent characteristics between data can be used. The inherent video alignment between background video and target video are considered in Ref. [22]. A smaller background matrix is generated through selecting short background video sequence related to the target video sequence, so that a lot of computing time is saved.

High resolution images can present more details, but they consume too much time and storage. On the one hand, the requirement for pixels is not high in practical monitoring application, so the images can be down-sampled, which can not only reduce the amount of data operation, but also reduce the noise in the spatial domain. Choosing the appropriate neighborhood size will play an important role in the results. If the neighborhood size is too large, the features concerned will be eliminated. While the improvement of algorithmic speed will be limited if the neighborhood size is too small. In order to compare the corresponding aligned frames, Ref. [10] used normalized cross-correlation (NCC) between the two images and applied multi-size NCC for different size detection objects. On the other hand, high resolution signals can be recovered from observed low-dimensional signals. High-res images can also be recovered from low-resolution images, that is, sparse coding migration [23]. Qin et al. proposed a fast and robust method for moving object detection in video sequences using sparse representation and code migration [8]. However, the model does not consider the online updating of the background dictionary, which results in unsatisfactory detection results for video sequences with global background changes.

3. Proposed Method

Each frame can be divided into foreground and background in the camera video. Background discrimination is similar to denoising, in which the foreground target can be treated as impulse noise. Assume that the camera video resolution is $p \times q$, the sequence of moving background is B_1, B_2, \dots, B_k , and the image sequence is X_1, X_2, \dots, X_m . The current video image can be represented as:

$$X = B + E \quad (6)$$

where B and E represent background image and moving object image respectively.

According to the characteristics of low rank and sparse for background matrix A and

foreground matrix E , the video matrix can be decomposed into low rank matrix and sparse matrix. Eq. (6) can be solved through principal component pursuit (PCP) :

$$\min_{B,E} \text{rank}(B) + \lambda \|E\|_0, \quad s.t. X = B + E \quad (7)$$

where λ is a non-negative parameter, $\|E\|_0$ is the l_0 norm of matrix E , and $\text{rank}(B)$ indicates rank function of matrix B .

Since the rank function and norm are non-convex, the matrix represented by Eq. (7) can be decomposed into a NP-hard problem, so the model needs to be modified on this basis. PCP method uses the kernel norm to approximate the rank of the matrix, and applies the l_1 norm of the matrix to constrain the sparsity, so that a solution model of convex optimization problem can be obtained:

$$\min_{B,E} \|B\|_* + \lambda \|E\|_1, \quad s.t. X = B + E \quad (8)$$

where $\|B\|_*$ is nuclear norm of matrix B and $\|E\|_1$ is l_1 norm of matrix E . The low rank background matrix and the sparse foreground matrix can be decomposed through the solution of the model.

Aiming to the target detection in video sequences, the initial background model can be obtained by using training samples, and then the background model can be updated by using dictionary learning method. Therefore, a minimum constraint function is constructed to initial modeling of background:

$$B_0 = \arg \min_{D,\alpha} \|B - D\alpha\|_2^2 + \mu \|\alpha\|_1 \quad (9)$$

where D is data dictionary to be trained. The initial training dictionary and the sparse coefficient vector of B can be obtained through K-SVD.

After the initial sparse representation model is obtained, the nearest neighbor frames ($X_n, X_{n+1}, \dots, X_{n+q}$) of the current time are extracted from the video sequence to be detected. Then K-SVD dictionary learning method is used to update the dictionary atoms (d_k) and sparse coefficients iteratively, and find the optimal data dictionary, so that the background image based on the dictionary can approximate the observation value of the adjacent frame image background optimally. The updated optimal background model can be obtained:

$$B = \arg \min_{D,\alpha} \|X - D\alpha\|_2^2 + \mu \|\alpha\|_1 \quad (10)$$

Then the determination of moving object can be obtained by background difference method. But if the update and calculation are needed every frame, the computation is too large and the real-time is difficult to meet the requirement. For many detection scenarios, the speed requirement of mobile camera is very low, and the speed is almost unchanged. The motion will follow a certain law of motion, so the foreground video can be calculated by recursive formula. While for many detection scenarios, the speed requirement of camera is very low, and the rate

is almost unchanged. So the foreground video can be calculated by recurrence formula according to the movement rules.

A set of data points are sampled from subspace set $S = U_{i=1}^J S^i$ uniformly. If there is enough sampling density, each sample can be expressed through a linear combination of other samples from the same subspace [20]. For slowly moving camera, continuous background video sequences can be considered as a series of subspace sets, which can be expressed as:

$$B_j = \sum_{i=1}^k B_i \alpha_i \quad (11)$$

Then the Eq. (8) can be transformed into:

$$\min_{W,E} \|B\|_1 + \lambda \|E\|_1 \quad s.t. \begin{cases} X = B + E \\ BW = B \\ W_{jj} = 0, \forall i. \end{cases} \quad (12)$$

Similarly, it can be considered that the video image can be recursively obtained from the previous video image. Combined with the dictionary sparse representation, the current video image sequence X_j can be represented as:

$$\hat{\alpha} = \arg \min_{B_i, \alpha_i} \left\| X_j - \sum_{i=1}^K B_i \alpha_i \right\|_2^2 + \left\| \sum \alpha_i \right\|_1 \quad (13)$$

If the background video is utilized as a dictionary to represent the current video image sequence, X_j can be sparse approximated through (B_1, B_2, \dots, B_k) . Then the moving object is determined by the difference method:

$$E = \begin{cases} 0 & \text{if } |I_j - B_j| < d \\ |I_j - B_j| & \text{else} \end{cases} \quad (14)$$

where d denotes set threshold. If the value is less than the threshold, it belongs to the background pixel, otherwise it is the foreground pixel.

Video sequences are taken as subspace set on the basis of low-rank sparse decomposition, and sparse approximation is applied to recursively represent the computational model. The processing schematic is shown in Fig. 2.

High-pixel cameras are often used in real-time surveillance to avoid introducing more noise in the acquisition process, and the high resolution images can present more details, but the time and storage consumption is too large. The improved model still has some requirements for time and memory storage. High and low resolution image blocks can be showed by the same sparse coefficient relative to their respective over-complete dictionaries [23]. Therefore, the above operations can be carried out in the down-sampling space, which can further reduce the computational complexity and storage requirements, and overcome the impact of large

anomaly areas. The videos are sampled down to subspace of different sizes. In order not to affect the final discrimination, the target image is mapped back on the sparse coefficient. The diagram of sparse approximation after down-sampling is shown in Fig. 3.

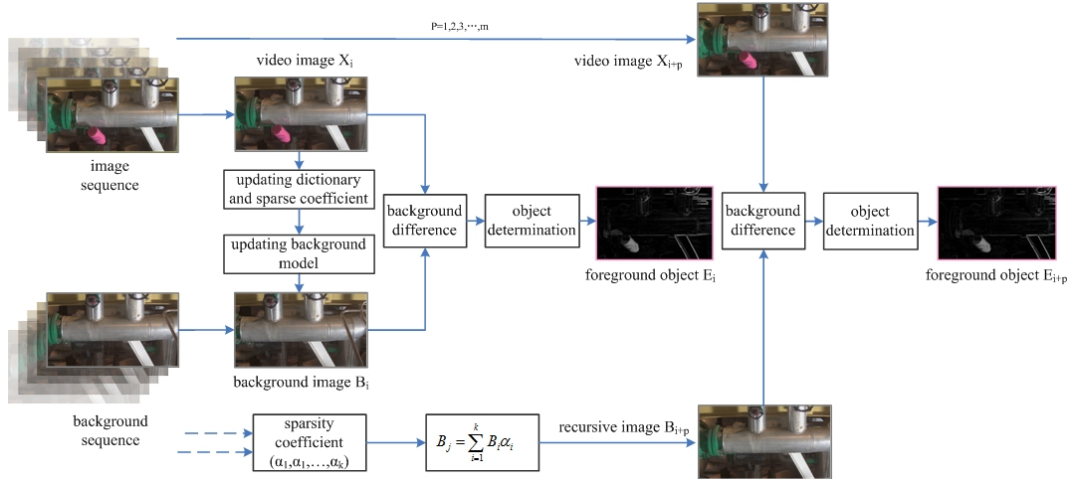


Fig. 2. Improved sparse approximation representation for target object detection

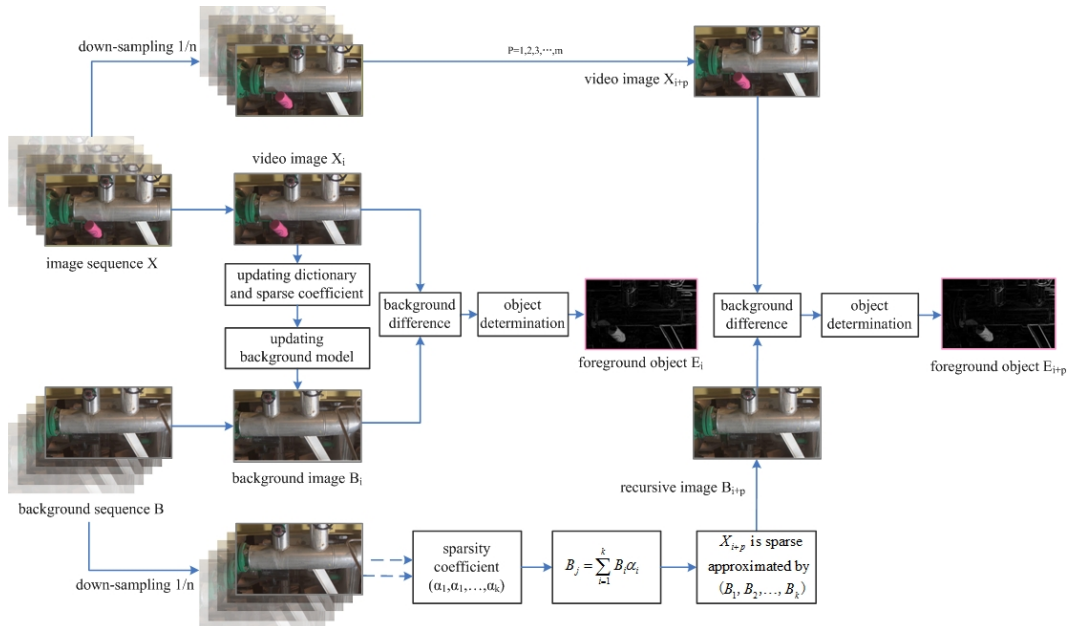


Fig. 3. Using sparse approximation after down-sampling to represent target object detection

4. Experimental Results and Analysis

4.1 VDAO Database for Object Detection

In order to evaluate the effectiveness and real-time of the proposed algorithm, VDAO database [24] is used for experimental comparison. The database contains 59 sets of target videos and corresponding background videos. Each target video contains 200 frames and a detection object. In addition, the dataset also annotates the position of the detected object in all video

frames in the form of providing the coordinates of the bounding box containing the object. All target videos contain nine different types of detection objects, such as camera box, bottle, shoe, towel and can, as shown in Fig. 4.



Fig. 4. Different types of detection objects are included in the database. The objects listed are camera box, bottle, shoe, towel and can

4.2 Comparison of Object Detection Methods

The performance of all methods is initially quantified by the following detection error rate indicators: TP (True positive) shows the correct number of foreground points, TN (True negatives) represents the correct number of background points, FP (False positives) indicates the number of foreground points for errors, FN (False negatives) represents the number of background points for errors, DIS is defined as the minimum distance from all operating points to the midpoint (1,0) of the TP × FP plane, PBC means the proportion of error points, Precision denotes the accuracy of foreground segmentation and Recall denotes the accuracy of foreground points.

$$PBC = 100 \times \frac{FN + FP}{TN + TP + FN + FP} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

We compared the performance of our proposed multi-SASL to the state-of-the-art mcRoSuRe-A [7], as well as the STC-mc [17], DAOMC [11], MCBS [12], and ADMULT [10] methods, and the result is shown in Table 1. Multi-SASL method outperforms other algorithms except ADMULT method. Although the result of ADMULT method is better than that of this method, multi-SASL method performs better in terms of 59 video results and running time (as shown in Table 2). The running time of this method is also improved by nearly 10 times.

Table 1. Comparison between multi-SASL method and STC-mc, DAOMC, MCBS, mcRoSuRe-A and ADMULT methods

Object	STC-mc [17]			DAOMC [11]			MCBS [12]			ADMULT [10]			mcRoSuRe-A[7]			multi-SASL		
	TP	FP	DIS	TP	FP	DIS	TP	FP	DIS	TP	FP	DIS	TP	FP	DIS	TP	FP	DIS
Towel	0.92	0.01	0.08	1.00	0.10	0.10	1.00	0.07	0.07	1.00	0.00	0.00	0.99	0.47	0.47	0.82	0.18	0.18
Shoe	0.90	0.04	0.11	1.00	0.04	0.04	1.00	0.28	0.28	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00
Pink bottle	0.99	0.13	0.13	1.00	1.00	1.00	1.00	0.96	0.96	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00
Camera box	1.00	0.03	0.03	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00
White jar	0.29	0.64	0.96	1.00	0.10	0.10	1.00	0.99	0.99	1.00	0.00	0.00	1.00	0.75	0.75	0.95	0.05	0.05

The detection result for a single video is shown in Fig. 5. Multi-SASL method and mcRoSuRe-A method are very similar to each other in detecting a type of object from a single video.

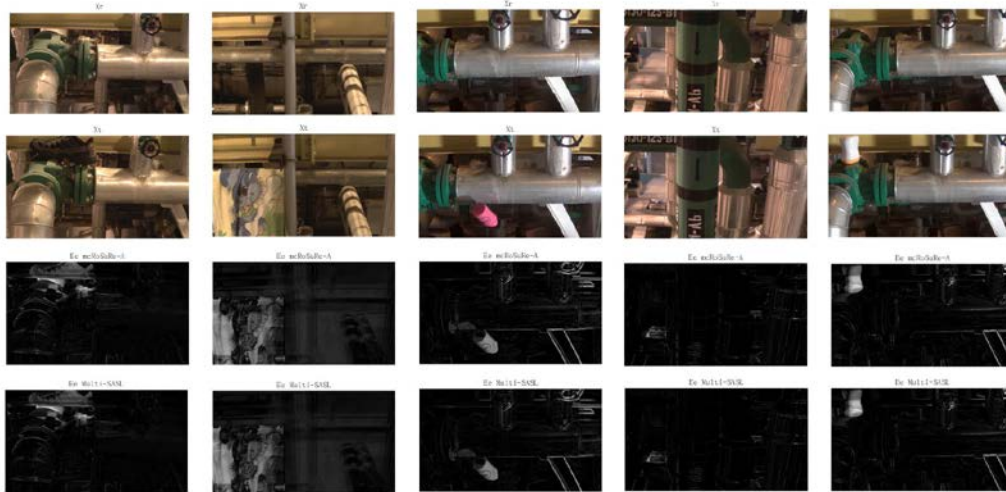


Fig. 5. Images from top to bottom are background video, target video, detection results of mcRoSuRe-A [7] and multi-SASL

In order to get the more comprehensive comparison of the results, the performance of the above algorithms is compared again on all VDAO videos in Table 2. Multi-SASL algorithm performs well in TP and TN values. In addition, the PBC value of this algorithm is the smallest, which indicates that the algorithm has fewer false pixels in moving target detection. And the Precision value is the highest. The higher the value, the more accurate the prospects of detection are.

Table 2. Comparison of detection results from STC-mc, DAOMC, MCBS, mcRoSuRe-A, ADMULT and multi-SASL on all videos

Method	TP	FP	TN	FN	PBC	Precision
STC-mc [17]	0.18	0.38	0.59	0.82	60.91	0.32
DAOMC [11]	0.83	0.43	0.54	0.17	30.46	0.66
MCBS [12]	0.89	0.84	0.02	0.11	51.08	0.51
ADMULT [10]	0.71	0.28	0.63	0.29	29.84	0.72
mcRoSuRe-A [7]	0.91	0.33	0.63	0.09	21.43	0.73
multi-SASL	0.91	0.24	0.72	0.08	16.84	0.79

4.3 Comparison of Time Performance

The algorithm reduces the complexity of the algorithm and improves the real-time response by using sparse approximation recursive formula and down-sampling. We evaluated the time with different down-sampling, assessed the impact on the accuracy, and compared the result with other advanced algorithms mentioned above. The running environment includes Intel Core i7-7800@ 3.50GHz processor and 31.1G memory.

The down-sampling sizes are 320 x 180, 80 x 45 and 64 x 36. While the processing size is too small to obtain enough data when 32×18 or smaller size is used, and this causes no change in the result. The processing time and detection results of all 59 videos are compared as shown in Table 3.

Table 3. Processing time (in seconds) and detection results of different down-sampling sizes

Size	TP	FP	TN	FN	Precision	PBC	Time
320×180	0.92	0.27	0.69	0.09	0.77	17.86	217
160×90	0.91	0.24	0.72	0.08	0.79	16.84	142
80×45	0.89	0.34	0.62	0.11	0.72	22.96	100
64×36	0.86	0.40	0.56	0.14	0.68	27.55	65

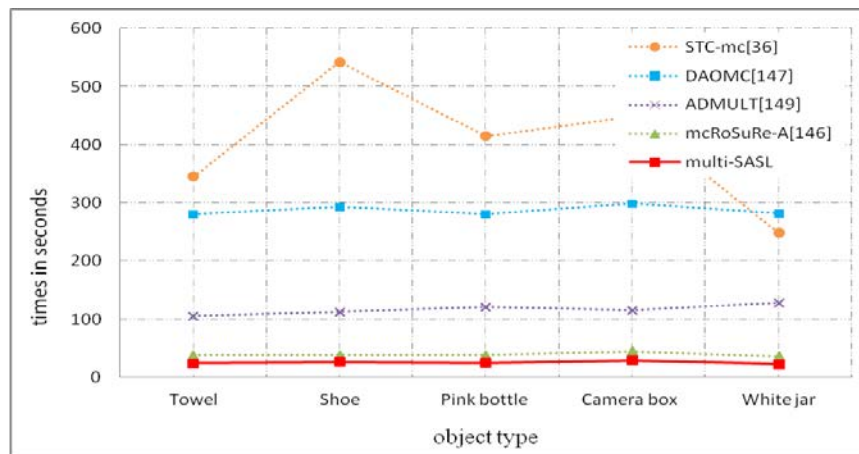
The processing time of 80×45 and 64×36 is very short, and the time of 160×90 is a little longer, but the overall performance index of 160×90 is the best.

The comparison between STC-mc [17], DAOMC [11], MCBS [12], mcRoSuRe-A [7] and ADMULT [10] and multi-SASL algorithm is shown in Table 4.

Table 4. Comparison of object detection processing time (in seconds) between STC-mc, DAOMC, MCBS, mcRoSuRe-A, ADMULT and Multi-SASL algorithm

Method	Towel	Shoe	Pink bottle	Camera box	White jar
STC-mc [17]	345	542	415	448	248
DAOMC [11]	280	293	280	299	282
MCBS [12]	50403	50427	50170	50238	49901
ADMULT [10]	105	112	121	115	128
mcRoSuRe-A [7]	38	38	38	45	36
multi-SASL	24	26	25	29	23

In order to better reflect the comparison, Fig. 6 shows the broken line comparison diagrams of various algorithms. While the running time of MCBS algorithm is removed from Fig. 6 to give a clearer comparison between other algorithms, because the operation time of MCBS algorithm is much worse than other algorithms. Experiment indicates our algorithm confirms detection precision while largely improves the detection efficiency.

**Fig. 6.** Comparison of running time between different algorithms

5. Conclusion

Aiming at the problem of object detection in the moving camera, this paper proposes a object detection algorithm based on sparse approximation recursion and sparse coding migration in

subspace. Firstly, continuous background video sequence is regarded as a series of subspace set. Low rank and sparse matrix decomposition is used to reduce the dimension of data, and dictionary is used to update the background model. Then background is used as dictionary to update the target video. Finally, background difference model is used to calculate the target object. In order to improve the real-time performance of the algorithm and reduce the requirement of memory storage, this paper proposes that the target model can be represented by sparse approximation and recursion, and the down-sampling process can be used simultaneously. At the same time, it can adapt to multi-scale target objects and overcome the impact of large anomaly areas. Through the experimental comparison and verification, our algorithm is more time efficiency and more space reduction than other object detection methods while keeping performance. Future work will focus on how to integrate a better camera motion model into the optimization constraint for low rank and sparse decomposition.

References

- [1] E. Bilgazyev, B. Efraty, S. Shah and I.A. Kakadiaris, "Sparse representation-based super resolution for face recognition at a distance," in *Proc. of 22nd British Machine Vision Conf.*, pp.1-11, August 29 - September 2, 2011. [Article \(CrossRef Link\)](#)
- [2] Y. Liu, Q. Zhang, J. Han and L. Wang, "Salient object detection employing robust sparse representation and local consistency," *Image and Vision Computing*, vol. 69, pp. 155-167, January, 2018. [Article \(CrossRef Link\)](#)
- [3] Z. Cong, X. G. Wang and K. C. Wai, "Background subtraction via robust dictionary learning," *EURASIP Journal on Image and Video Processing*, vol. 1, pp. 1-12, January, 2011.
- [4] N. Oliver, B. Rosario and A. Pentland, "A bayesian computer vision system for modeling human interactions," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831-843, August, 2000. [Article \(CrossRef Link\)](#)
- [5] S. Javed, S. K. Jung, A. Mahmood and T. Bouwmans, "Motion-aware graph regularized RPCA for background modeling of complex scenes," in *Proc. of 23rd Int. Conf. on Pattern Recognition*, pp. 120-125, December 4-8, 2016. [Article \(CrossRef Link\)](#)
- [6] W. Hu, Y. Yang, W. Zhang and Y. Xie, "Moving object detection using tensor-based low-rank and saliently fused-sparse decomposition," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp.724-737, February, 2017. [Article \(CrossRef Link\)](#)
- [7] L. Thomaz, E. Jardim, A. da Silva, E. da Silva, S. Netto and H. Krim, "Anomaly detection in moving-camera video sequences using principal subspace analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 3, pp. 1003-1015, March, 2018. [Article \(CrossRef Link\)](#)
- [8] X. Qin, G. Yuan, C. Li and X. Zhang, "An approach to fast and robust detecting of moving target in video sequences," *Acta Electronica Sinica*, vol. 45, no. 10, pp. 2355-2361, October, 2017.
- [9] H. K. Galoogahi, A. Fagg, C. Huang, D. Ramanan and S. Lucey, "Need for speed: a benchmark for higher frame rate object tracking," in *Proc. of IEEE Int. Conf. on Computer Vision*, pp. 1134-1143, October 22-29, 2017. [Article \(CrossRef Link\)](#)
- [10] G. H. F. de Carvalho, L. A. Thomaz, A. F. da Silva, E. A. B. da Silva and S. L. Netto, "Anomaly detection with a moving camera using multiscale video analysis," *Multidimensional Systems and Signal Processing*, vol. 30, no. 1, pp. 311-342, January, 2019. [Article \(CrossRef Link\)](#)
- [11] H. Kong, J. Audibert and J. Ponce, "Detecting abandoned objects with a moving camera," *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 2201-2210, August, 2010. [Article \(CrossRef Link\)](#)

- [12] H. Mukojima, D. Deguchi, Y. Kawanish, I. Ide, H. Murase, M. Ukai, N. Nagamine and R. Nakasone, "Moving camera background-subtraction for obstacle detection on railway tracks," in *Proc. of 23rd IEEE Int. Conf. on Image Processing*, pp. 3967-3971, September 25-28, 2016. [Article \(CrossRef Link\)](#)
- [13] Y. Zhou and S. Maskell, "Moving object detection using background subtraction for a moving camera with pronounced parallax," in *Proc. of 2017 Symposium on Sensor Data Fusion: Trends, Solutions, Applications*, pp.1-6, October 10-12, 2017. [Article \(CrossRef Link\)](#)
- [14] L.Gong, M.Yu and T.Gordon, "Online codebook modelling based background subtraction with a moving camera," in *Proc. of 3rd Int. Conf. on Frontiers of Signal Processing*, pp. 136-140, September 6-8, 2017. [Article \(CrossRef Link\)](#)
- [15] J. C. Yang, J. Wright, T. Huang and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.1-8, June 24-26, 2008. [Article \(CrossRef Link\)](#)
- [16] D. L. Donoho and X. M. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transaction on Information Theory*, vol.47, no.7, pp. 2845-2862, 2001. [Article \(CrossRef Link\)](#)
- [17] M. T. Nakahata, L. A. Thomaz, A. F. da Silva, E. A. B. da Silva and S. L. Netto, "Anomaly detection with a moving camera using spatio-temporal codebooks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 3, pp. 1003-1015, 2018. [Article \(CrossRef Link\)](#)
- [18] H. X. Li, C. H. Shen and Q. F. Shi, "Real-time visual tracking using compressive sensing," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1305-1312, June 20-25, 2011. [Article \(CrossRef Link\)](#)
- [19] X. Zhou, C. Yang and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 597-610, march, 2013. [Article \(CrossRef Link\)](#)
- [20] X. Bian and H. Krim, "Bi-sparsity pursuit for robust subspace recovery," in *Proc. of IEEE Int. Conf. on Image Processing*, pp. 3535-3539, September 27-30, 2015. [Article \(CrossRef Link\)](#)
- [21] E. Jardim, X. Bian, E. A. B. da Silva, S. L. Netto and H. Krim, "On the detection of abandoned objects with a moving camera using robust subspace recovery and sparse representation," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Process*, pp. 1295-1299, April 19-24, 2015. [Article \(CrossRef Link\)](#)
- [22] L. A. Thomaz, A. F. da Silva, E. A. B. da Silva, S. L. Netto and H. Krim, "Detection of abandoned objects using robust subspace recovery with intrinsic video alignment," in *Proc. of IEEE Int. Conf. Symposium on Circuits and Systems*, pp. 1-4, May 28-31, 2017. [Article \(CrossRef Link\)](#)
- [23] J. Yang, J. Wright, T. S. Huang and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861-2873, November, 2010. [Article \(CrossRef Link\)](#)
- [24] VDAO, "Video database of abandoned objects in a cluttered industrial environment," 2016. [Online] <http://www.smt.ufrj.br/~tvdigital/database/objects>.



Shufang Li is associate professor in Hebei University of Environmental Engineering, China. Her current research interests include crowd counting, pattern recognition. Li received her M.S. degrees from East China University of Technology, China, in 2007. She is currently pursuing the Ph. D. degree in School of Information Science and Engineering, Yanshan University, China. (Email: lishufang116003@sina.com)



Zhengping Hu received the B.S. and M.S. degrees in Radio Technology and Circuit and System from YanShan University, China, in 1996 and 1999, respectively. He received his Ph.D. degree in Information and Communication Engineering from Harbin Institute of Technology, Harbin, China, in 2007. He is a professor, Ph.D. supervisor, and Dean of Department of Electronic and Communication of YanShan University. His main research area covers the theories and algorithms related to pattern recognition, one -class classifier and image processing. (Email: hzp_ysu@163.com)



Mengyao Zhao received the B.S. degree in Electronic Information Engineering from YanShan University, China, in 2017. She is currently a master student majoring in electronics science and technology in YanShan University, Qinhuangdao, China. Her current research interest is in video anomaly detection. (Email: zhaomengyao0826@126.com)